



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Merrow et al.
Serial No. 09/945,282
Filing Date: August 31, 2001
Title: SPEECH RECOGNITION METHOD AND SYSTEM FOR DETERMINING THE
STATUS OF AN ANSWERED TELEPHONE DURING THE COURSE OF AN OUTBOUND
TELEPHONE CALL
Art Unit: 2697
Examiner: Elahee, Md S
Docket No. 57622-048 (ELZK-8)

CERTIFICATE OF MAILING (37 C.F.R. 1.8(a))

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to Mail Stop Non-Fee Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date indicated below.

Date: February 24, 2005

Cynthia Joseph
Cynthia Joseph

Mail Stop Fee Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF PROFESSOR LUIS RODRIGUES
UNDER 37 C.F.R. § 1.132

I, Luis Rodrigues, being over the age of 18 years, and competent to make this declaration, declares as follows:

1. I currently reside at 272 Oakdale Crescent, Beaconsfield, Quebec, H9W 1X2, Canada.
2. I received my PhD in Aeronautics and Astronautics from Stanford University, June 2002. During my PhD, I was a visiting student at MIT from June 2000 to January 2002. I received my Master of Science in Electrical Engineering and Computers, March 2000 from Instituto Superior Técnico - Universidade Técnica de Lisboa. I received my "Licenciatura" Degree in Electrical Engineering and Computers in September 1995 from Instituto Superior Técnico, Universidade Técnica de Lisboa.

3. During my degree in Electrical Engineering and my Master's I have had a significant exposure to Signal Processing in courses such as Signals Theory, Digital Signal Processing,

Design of Digital Signal Processing Control Systems (including a project on Word Recognition using Markov Chains), Image Processing and Robotic Vision. During my PhD at Stanford I have also continued working with signals from satellites, namely on a project of automatic docking of a boat using GPS measurements. After completing my PhD I have worked as a consultant for Eliza Corporation on Articulatory Speech Production Models and Speaker Recognition from February 2002 to August 2002. I then joined Ydreams and worked as project manager on a virtual fly-over a rural landscape project from September to December 2002. The objective was to use 3D interactive tools and geographical information systems (GIS) to simulate an aircraft fly-over a rural landscape with precise geographic information for spatial plan preparation.

4. Since January 2003 I am Assistant Professor at Concordia University in the Department of Mechanical and Industrial Engineering.

5. My research interests lie primarily in the areas of guidance, navigation and control systems design with applications to aerospace, convex optimization methods for control systems, control with limited information, control of biological systems (including the vocal tract), speech production models and speech recognition using system theoretical methods.

6. I have reviewed the above-identified application, the Office action dated January 13, 2004, the references cited and applied in that Office action (the Matthews et al., Bartholomew et al. , Miner et al., Szlam et al. and Brown et al. patents), and Amendment B submitted to the USPTO on July 13, 2004, including the current version of the claims as submitted as a part of Amendment B, in response to that Office action.

7. My research activities have provided me with significant expertise and exposure in the field of digital signal processing, and its applications in such areas as speech recognition and voice verification.

8. Applications for computer processing of speech fall into two main areas:

a. PC or server based processing, whereby a user speaks into a microphone that is connected directly to a personal computer, server, or other hardware system'

b. Over the telephone processing, whereby a user speaks into a telephone, with the voice signals transmitted, at least in part, over the public telephone network to a server that performs the analysis.

9. The focus of the above-identified patent application and the current version of the claims of the application is in “over the telephone” technology, so I will focus my comments accordingly on the second category (8b.) above.

10. There are two fundamentally different types of technology in speech processing: speaker-dependent and speaker-independent. Speaker-dependent technology can be used to provide speaker verification/identification. Voice verification involves verifying the identity of a given caller (or recipient of a call) by comparing an utterance to a pre-defined (i.e., stored) voice pattern template. Typically, users must “train” or “register” with such a system by calling into it and recording a password or phrase several times. This set of pre-defined, sample utterances is used to determine the relevant voice characteristics against which to compare future utterances – to verify the identity of a given speaker (as opposed to determining the semantic meaning of what the speaker said). The terms “voice authentication”, “speaker verification,” and similar nomenclature provided below are understood by those skilled in the art to refer to this speaker-dependent branch of the technology.

11. Speaker-dependent technology can also be used to recognize utterances of a particular user who has trained or registered on such a system (usually over an extended period of time and many utterances). Examples of such speaker-dependent technology is described in the patents issued to Matthews et al. and Bartholomew et al., both cited in the Office action of January 13, 2004. The terms “dictation” and “transcription” are understood by those skilled in the art to refer to this second speaker-dependent branch of the field.

12. Speaker-independent technology is used to provide over the telephone speech recognition, whereby the system determines the semantic meaning of an utterance from the speaker (as opposed to the speaker’s identity), without any pre-defined, stored voice pattern template. In the context of telephony, the term “speech recognition” (SR) is understood by those skilled in the art to refer to this speaker-independent branch of the field. SR is the focus of the above-identified patent application.

13. The limitation of having a pre-defined voice pattern template is not required, nor is it suggested or inherently necessary in the patent disclosure and pending claims of the above-identified application. This is not surprising, given the accepted industry nomenclature. I believe that those skilled in the art would understand that, in the context of outbound calling, the term “speech recognition” refers to the speaker-independent analysis of voice data to determine

semantic meaning (e.g., to determine call control or application branching in an outbound telephone call).

I believe that the following nomenclature is commonly accepted in the speech recognition industry:

Industry Nomenclature

Speaker-Dependent

Voice Verification

Voice Authentication

Voice Recognition

Speaker Verification

Speaker Identification/Authentication

Speech-Based Identification/Authentication

Voice-Based Identification/Authentication

Voice Biometrics

Transcription

Dictation

Speaker-Independent

Telephonic Speech Recognition

Telephonic Automatic Speech
Recognition

14. Furthermore, given the nature/purpose of the claims of the above-identified application, i.e., for outbound calling, one skilled in the art would realize that a pre-defined voice pattern template is not required (indeed, without such limitation being specified). That is, outbound calls would be placed to households containing individuals who had not registered on the system, and furthermore no system can control who picks up the phone at any given household (such that only someone who had registered on the system would answer said outbound call). Therefore, the limitation that such a template is not required for the speech recognition analysis of the claimed invention of the above-identified application should not need to be specified.

15. As to the claims of the above-identified application, it is understood by those skilled in the art that speech recognition technology is extremely context-dependent. Those skilled in the art realize that substituting SR in a new situation or application is unlikely to be successful, without a complete analysis, understanding, and modeling of the new situation. It is the

modeling of each such situation that enables the technology to function. Indeed, the modeling of such distinct, context-dependent interactions is currently recognized as a major challenge for the speech recognition technology industry, with not even the smallest of combinations or adaptations being taken for granted.

16. In particular, inbound calls are, surprisingly, completely different from outbound calls.

17. A person making an inbound call has a definite purpose, which can be exploited by the automated system (by narrowing the recognition search to words/grammars that are expected in the caller's purpose/context). An outbound call to a household must, on the other hand (by definition), be unexpected, does not have a strong context to define potential user utterances, and must cope with the full range of human behavior - of not only the target person, but possible family members and so on. Hence, the specific modeling of outbound calling states provided in various claims of the above-identified application. It is my opinion that it would not have been obvious to one skilled in the art at the time of the filing of the above-identified application, to combine a series of differing inbound and speaker verification functions described in the Matthews et al., Bartholomew et al. , Miner et al., Szlam et al. and Brown et al. to develop an outbound model for SR, as defined by the claims of the above-identified application.

18. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

OCTOBER 28, 2004
Date

Luis Rodriguez
Luis Rodriguez